RESEARCH ARTICLE

Genomic architecture of phenotypic divergence between two hybridizing plant species along an elevational gradient.

ABSTRACT

Knowledge of the genetic basis of phenotypic divergence between species and how such divergence is caused and maintained is crucial to an understanding of speciation and the generation of biodiversity. The hybrid zone between *Senecio aethnensis* and *S. chrysanthemifolius* on Mount Etna, Sicily, provides a well-studied example of species divergence in response to conditions at different elevations, despite hybridization and gene flow. Here, we investigate the genetic architecture of divergence between these two species using a combination of quantitative trait locus (QTL) mapping and genetic differentiation measures based on genetic marker analysis. A QTL architecture characterized by physical QTL clustering, epistatic interactions between QTLs, and pleiotropy was identified, and is consistent with the presence of divergent QTL complexes resistant to gene flow. A role for divergent selection between species was indicated by significant negative associations between levels of interspecific genetic differentiation at mapped marker gene loci and map distance from QTLs and hybrid incompatibility loci. Within-species selection contributing to interspecific differentiation was evidenced by negative associations between interspecific genetic differentiation and genetic diversity within species. These results show that the two *Senecio* species, while subject to gene flow, maintain divergent genomic regions consistent with local selection within species and selection

against hybrids between species which, in turn, contribute to the maintenance of their distinct phenotypic differences.

KEYWORDS: Genetic differentiation; hybridization; phenotypic divergence; QTL architecture; QTL interactions; selection; speciation

INTRODUCTION

Speciation commonly proceeds through genetic divergence between populations that ultimately become reproductively isolated from each other due to intrinsic and/or extrinsic breeding barriers (Orr and Turelli 2001; Coyne and Orr 2004; Smadja and Butlin 2011; Nosil and Feder 2012). Phenotypic trait divergence usually accompanies this process, often as a result of adaptation to different environments (Nosil 2012). Understanding how phenotypic trait divergence evolves between populations and is maintained between hybridizing species requires knowledge of the genetic basis of divergent traits and how selection acts on genes controlling these traits (Rieseberg *et al.* 2003; Lexer *et al.* 2005; Nosil *et al.* 2009; Nosil and Feder 2012).

Quantitative trait locus (QTL) analysis is a powerful way of analysing the genetic basis of divergent traits between species (Rieseberg *et al.* 2003; Lexer *et al.* 2005; Bouck *et al.* 2007; Taylor *et al.* 2012; Rogers *et al.* 2013). It involves determination of the number and primary effects of QTLs, their genomic locations, the interactions between them (epistasis), and their effects across multiple traits (pleiotropy). The QTL architecture of divergent traits revealed by such analysis is likely to be shaped by divergent selection acting against relatively unfit recombinant hybrid phenotypes (Bierne *et al.* 2011; Servedio *et al.* 2011; Abbott *et al.* 2013; Yeaman 2013), especially where divergence between species occurs in the presence of interspecific gene flow (Via and West 2008; Nosil *et al.* 2009; Yeaman and Whitlock 2011; Yeaman 2013). This selective scenario could favour the evolution of QTL hotspots, epistasis, and pleiotropy as effective means of preserving local adaptation

59 despite gene flow (Whiteley *et al.* 2008; Gagnaire *et al.* 2013; Lindtke and
60 Buerkle 2015). Alternatively, recombination and break-up of QTL complexes
61 could be reduced by close physical proximity of QTLs (Yeaman and Whitlock
62 2011; Jones *et al.* 2012; Yeaman, 2103) or recombination 'coldspots' such as
63 near centromeres or chromosomal rearrangements (Turner *et al.* 2005;
64 Kirkpatrick and Barton 2006; Lowry and Willis 2010; Twyford and Friedman
65 2015).

66 Complementary insights into the relationship between QTL architecture
67 and divergent selection can be obtained by investigating genetic diversity and
68 differentiation among mapped molecular marker loci (Rogers and Bernatchez
69 2007; Stinchcombe and Hoekstra 2008; Gompert *et al.* 2012; Renaut *et al.*
70 2012; Strasburg *et al.* 2012, Cruikshank and Hahn 2014). Heterogeneous
71 differentiation across the genome is expected to result from divergent
72 selection in the presence of gene flow (Wu 2001; Feder and Nosil 2010) and
73 has been reported in several studies of ecologically divergent pairs of taxa
74 (Turner *et al.* 2005; Rogers and Bernatchez 2007; Via and West 2008).
75 However, such patterns of differentiation can be highly dependent on the
76 biology and demographic histories of the focal taxa (Jones *et al.* 2012; Renaut
77 *et al.* 2012), and their assessment must take account of genetic diversity both
78 within and between focal taxa (Cruikshank and Hahn 2014).

79 Here we present a quantitative genetic analysis of divergent traits
80 between two diploid (2n = 20), short-lived perennial, self-incompatible,
81 herbaceous species of *Senecio* (Asteraceae), *S. aethnensis* and *S.*
82 *chrysanthemifolius*, which grow at elevations above 2000 m and below 1000
83 m, respectively, on Mount Etna, Sicily. Whereas *S. aethnensis* produces large

flower heads (capitula) and fruits, and entire (spathulate) leaves, *S. chrysanthemifolius* has smaller flower heads and fruits, and highly dissected (pinnatisect) leaves. The two species hybridize and form a hybrid zone at intermediate elevations on Mount Etna (James and Abbott 2005; Abbott and Brennan 2014). Although connected by hybrid populations, some barriers to interspecific gene flow are apparent in the field. For example, flowering times only partially overlap, with *S. chrysanthemifolius* flowering six weeks earlier (April to June) than *S. aethnensis* (July to September) (authors' personal observation). A previous analysis of the hybrid zone showed that leaf shape, flower head structure, and fruit structure exhibited steeper clines and/or shifts in cline position relative to a molecular genetic cline (Brennan *et al.* 2009). This was attributed to both intrinsic and extrinsic environmental selection against hybrids.

An improved understanding of the level of genetic divergence between the two species and the importance of selection in driving genomic divergence recently came from a comparison of their transcriptomes (Chapman *et al.* 2013). This showed that genome-wide genetic differentiation between the species was low, with only 2.25% of 8,854 loci tested having been subject to divergent selection. Genetic maps for the two *Senecio* species based on segregation of molecular markers in F_2 mapping families (Brennan *et al.* 2014; Chapman *et al.* 2015) indicated that large genomic rearrangements were not a cause of reduced fitness in hybrids. However, many markers (~27% of 127 maker loci tested, Brennan *et al.* 2014) exhibited significant transmission ratio distortion (TRD) in the F_2 family and clusters of transmission ratio distortion loci (TRDLs) were distributed across multiple linkage groups. This frequency

of TRD was similar to that found in other crossing studies involving distinct "species" (e.g. 49 and 33 % in *Mimulus* and *Iris*, respectively, Fishman *et al.* 2001; Taylor *et al.* 2012). Such extensive genomic incompatibility between the two species would be expected to impact the genetic structure of the hybrid zone on Mount Etna by limiting interspecific gene flow and promoting divergence across the genome. Chapman *et al.* (2015) further showed that loci exhibiting significant sequence or expression differentiation between the two species had a clustered distribution when placed on the map and several QTLs for species phenotypic differences coincided with these regions.

Here we investigate further the genetic architecture of phenotypic trait differences and associated divergent selection acting on *S. aethnensis* and *S. chrysanthemifolius* by performing a QTL analysis of multiple quantitative traits that distinguish the two species. Our analysis examined additional traits and a larger mapping family relative to the recent study by Chapman *et al.* (2015), albeit with a reduced number of molecular marker loci. Our study aimed to determine the number and genomic locations of QTLs of relatively large effect controlling phenotypic differences and the extent of epistatic and pleiotropic effects of QTLs that could limit introgression between the two species in the wild. We also conducted genetic differentiation outlier tests on mapped molecular markers in the two species to identify loci under divergent selection and test for associations between outlier loci and QTLs. In addition, we tested if previously identified hybrid incompatibilities are associated with either QTLs for species differences or highly divergent loci as would be expected under divergent selection.

METHODS

Samples

An F_2 mapping family (F_2AC) of a reciprocal cross between two cross-compatible F_1 progeny derived from a reciprocal cross between *S. aethnensis* (A) and *S. chrysanthemifolius* (C) was produced as described in Brennan *et al.* (2014) and used for QTL analysis. This family consisted of 100 individuals of known parental cytotype. For tests of selection based on genetic differentiation, seed was collected from two wild populations of *S. aethnensis* and three of *S. chrysanthemifolius* representing the elevational extremes of each species' range and also the source locations of the mapping family parents (NIC1 and PIC1, Table S1) [**see SUPPORTING INFORMATION**]. Forty-two plants of each species, each representing a separately sampled maternal individual, were raised from this seed in a glasshouse at the same time and under the same conditions as F_2AC individuals.

Phenotype measurement

Twenty-five traits were measured on F_2AC parents and progeny, and also wild sampled individuals (see Brennan *et al.* 2009 for a description of traits measured). Extreme outlier values > 3 standard deviations from the mean were removed from the datasets for progeny and wild samples of each species prior to analysis. Trait summary statistics were calculated and comparisons between wild sampled *S. aethnensis*, wild sampled *S. chrysanthemifolius*, and the F_2AC mapping family, were made using one-way analyses of variance and Mann Whitney tests. Three traits: capitulum length, ray floret number, and selfing rate were dropped from further analysis after preliminary data exploration found that they showed extreme distributions that could not be satisfactorily resolved with data transformations. Remaining trait

159 measurements were not transformed to become normally distributed before
160 QTL analysis because (i) the expected density distributions of traits with
161 additive effects contributed by multiple loci are not necessarily normally
162 distributed, (ii) the significance of QTL logarithm of odds scores (LOD) can be
163 adequately assessed with data permutation, and (iii) estimated sizes of QTL
164 effects are more directly interpretable based on untransformed data (Churchill
165 and Doerge 1994). Cross direction did not significantly influence any trait
166 mean, so this was not required as a cofactor for QTL analysis. Independence
167 between measured traits was examined using paired-trait Spearman
168 correlations and tests of their significance were performed separately for wild
169 sampled *S. aethnensis*, wild sampled *S. chrysanthemifolius*, and the F_2AC
170 mapping family progeny leading to a subset of 13 highly independent traits
171 being retained for QTL analysis. All tests were performed using R v2.13
172 software (R Development Core Team 2011).

173 **DNA isolation and genotyping**

174 DNA was extracted from each plant using the method described by
175 Brennan *et al.* (2009). Plants were genotyped across 127 marker loci
176 comprising 77 AFLPs, eight SSRs, and 42 EST-SSRs and indel molecular
177 markers as described by Brennan *et al.* (2014). For about 10% of plants
178 (randomly chosen), two independent DNA extracts were made to test for
179 genotyping reliability.

180 **Genetic mapping**

181 A genetic map was constructed from the segregation of genetic
182 markers in the F_2AC mapping family as described in Brennan *et al.* (2014)
183 and supplementary information. Genotype uncertainty due to scoring of

dominant markers was accounted for by using the MapMaker genotype classes C (not a homozygote for the first parental allele) and D (not a homozygote for the second parental allele; Lander *et al.* 1993). The genetic map comprised 10 independent linkage groups with a total length of approximately 400 cM. Transmission ratio distortion affected ~27% of mapped markers that were clustered into nine transmission ratio distortion loci (TRDLs). Sixty-five mapped loci were included in the QTL analysis after removing 39 loci that did not show F_2-like allelic segregation (i.e. each parent had an allele in common) and 23 loci that were located less than 0.5 cM from the nearest neighbouring marker and which therefore added little extra QTL mapping power.

Quantitative trait locus mapping and analysis

We analysed the data in the form of individual differences from the combined species mean, with the sign altered so that individuals that were more similar to *S. aethnensis* or *S. chrysanthemifolius* mean values were positive and negative, respectively. This data transformation preserved effect sizes in original units, but had the added advantage of standardizing effect directions according to parental species across all traits. Comparisons with untransformed data showed that LOD scores (base ten logarithm of odds) were largely unaffected by the transformation. Multiple interval mapping (MIM) was used to identify QTLs because this method has the advantage of simultaneously accounting for multiple QTLs and their interactions (Kao and Zeng 1999). MIM was performed with QTL cartographer v2.5.10 (Wang *et al.* 2011) using forward regression with a scanning interval of 3 cM and Bayesian Information Criterion (BIC-M0) model selection to determine the inclusion of

extra QTL or QTL interaction parameters. Initial MIM models were then refined by testing indicated QTLs for significance according to BIC and adding additional QTLs until no further significant model improvement was achieved. Epistatic QTL interactions were also included if BIC was significantly improved. For comparison with MIM, composite interval mapping (CIM; Zeng 1994), a widely used QTL mapping method, was also performed and results obtained from this analysis, which did not differ greatly from those obtained with MIM, are presented in Supplementary Information. The potential for transmission distortion loci (TRDLs) to influence the QTL results was tested using Spearman rank correlation tests of marker distance to nearest QTL peak against marker Chi-square test values for segregation distortion of genotypes, heterozygotes, and parental alleles.

Multiple trait composite interval mapping (MtCIM) simultaneously analyses multiple trait data and can distinguish between linked QTLs and a single QTL affecting more than one trait through pleiotropy (MtCIM; Jiang and Zeng 1995). MtCIM analysis was performed using a scanning interval of 3 cM and automatic model selection using forward regression with five cofactor loci outside the test interval window of 10 cM. Significance of QTL LOD scores was tested with 1000 permutations of trait values (Churchill and Doerge 1994). A complementary test of the extent to which QTLs for different traits occupied the same genomic regions applied the “sampling without replacement” method (Moyle and Graham 2006; Paterson 2002). Because the traits examined in this QTL dataset were selected to minimize covariance between them, spurious patterns of QTL coincidence generated by covariance were also assumed to be minimized, avoiding the need for

additional statistical correction (Breitling *et al.* 2008). To perform the "sampling without replacement" test, the genetic map was divided into smaller intervals of equal size corresponding to the mean QTL 2-LOD cM confidence interval of 16.5 cM with intervals chosen to be centred over each linkage group. This level of subdivision of the genetic map generates an optimal proportion of intervals occupied by a QTL for the purposes of this test (Patterson 2002), but the effect of using smaller interval sizes was also tested by repeating the test with 2, 4, 6, 8, 10, 12, and 14 cM interval sizes. A binary matrix describing the presence or absence of QTLs for each trait within intervals was constructed and for each pair of traits, the probability of coincidence (p) was tested according to:

$$p = \binom{l}{m}\binom{n-l}{s-m} / \binom{n}{s}$$

where *n* is the number of intervals compared, *l* and *s* are the number of QTL intervals present in the samples with larger and smaller QTL counts, respectively, and *m* is the number of paired QTL interval matches present. To test whether QTL coincidence was greater than the null hypothesis of a random distribution of QTLs across the genetic map, the observed mean probability of QTL coincidence across paired-trait comparisons was compared against the distribution obtained from 1000 random permutations of QTL locations. The coincidence between TRDLs and QTLs was also investigated by including TRDL data in this analysis.

Genetic diversity analysis

Summary population genetic statistics were estimated for all mapped markers genotyped in wild samples of *S. aethnensis* and *S. chrysanthemifolius*. The population genetics software used included: Arlequin

(Excoffier and Lischer 2010), GenAlEx v6.1 (Peakall and Smouse 2006) and HPrare (Kalinowsky 2005). The estimated statistics for AFLP and other dominantly-scored markers were band presence frequency (*p*; assuming Hardy-Weinberg equilibrium), effective number of alleles (*Ne*), unbiased heterozygosity (*UHe*), allelic richness (*Ar*), private allelic richness (*pAr*), genetic differentiation among species (F_{ST}) and genotypic differentiation (Φ_{PT}). The same statistics, excluding *p* but including the minor allele frequency (*MAF*) and inbreeding coefficient (F_{IS}), were calculated for codominantly scored markers.

Patterns of differentiation across loci were investigated to detect both strongly and weakly differentiated outlier loci using BayeScan (Foll and Gaggiotti 2008), which employs Bayesian methods to estimate locus-specific differentiation and to evaluate its probability relative to population-level differentiation. Default starting parameter settings were used, except for a Monte Carlo Markov Chain (MCMC) size of 10000, thinning interval of 50, ten pilot runs of 10000, and an additional burn in of 100000. Outlier loci were identified based on $\log_{10}$ Bayes Factors (BF) values greater than one. Outlier analysis was performed with individuals classified according to both species and population. Initial runs suggested that loci with very low MAF were over-represented among outliers. To overcome this problem, only those loci with MAF greater than 0.05 were included in final differentiation analyses, which were conducted separately on datasets comprising 64 codominant loci and 132 dominant loci.

The presence of “genomic islands” of divergence was investigated by testing the genomic clustering of outlier markers with binomial tests that the

observed number of neighbouring pairs of significantly selected loci was greater than the expected number of neighbouring paired selected loci given by the square of the observed frequency of selected loci. Genetic differentiation, measured as both F_{ST} and Φ_{PT}, was tested for an association with the genetic map distance to the nearest QTL peak and the nearest TRDL peak using Spearman rank correlation tests. Genetic differentiation was further tested for associations with local recombination rate, measured as the genetic map distance to the nearest mapped marker, and with genetic diversity within species, measured as each of *UHe* and *Ar* and *MAF* using Spearman rank correlation tests. Marker loci on linkage groups without QTLs were assigned large QTL distance values of 50 cM in order to include them as part of these association tests.

RESULTS

Quantitative trait locus mapping and analysis

The two parent species, *S. aethnensis* and *S. chrysanthemifolius*, differed significantly for 22 of the 25 traits. The exceptions were flowering time, leaf number, and selfed seed-set (Traits 1, 3 and 18, Table S2, Figure S1) [**see SUPPORTING INFORMATION**]. We surmise that the lack of flowering time difference in the glasshouse compared to field observations reflects the importance of environmental conditions for the expression of this trait. For example, suitable growing conditions at the onset of spring start later in higher elevation *S. aethnensis* habitat than lower elevation *S. chrysanthemifolius* habitat. In summary, *S. aethnensis* differed from *S. chrysanthemifolius* in being shorter and less branched, possessing smaller, less dissected leaves (i.e. having entire or slightly lobed edges), and fewer but

309 larger capitula that produced larger seed. Significant differences between the
310 mean of the F_2AC family and those of one or both parent species were also
311 evident for all traits apart from pollen viability and selfed seed-set (Traits 16,
312 18, Table S2) [**see SUPPORTING INFORMATION**]. The means of the F_2AC
313 family for all traits were neither significantly higher nor lower than the means
314 of both parents. Paired trait correlations are summarized in Table S3. Overall,
315 4.3, 2.7, and 11% of pairs of traits were significantly correlated after correction
316 for multiple testing among progeny of wild sampled *S. aethnensis*, wild
317 sampled *S. chrysanthemifolius*, and the F_2AC mapping family, respectively.
318 Instances of non-independence between traits were reduced by dropping
319 highly correlated traits and traits used to calculate compound characters,
320 leaving a subset of 13 independent traits for QTL analysis.
321 Significant QTLs for each trait were detected and characterized by
322 LOD score, map position, two LOD confidence intervals, size of additive,
323 dominance, and epistatic effects, and percentage variance explained (PVE). A
324 total of 29 significant QTLs were detected across the 13 traits examined with
325 mean QTL effect size of 15 % (Table 1, Figure 1). QTLs were distributed
326 across all major linkage groups except AC3 and AC6, with one to five QTLs
327 detected for each trait (Figure 1). The mean PVE of all identified QTLs per
328 trait was 33.5 % (range 10.0 – 69.8 %).
329 Four pairs of QTLs exhibited significant epistatic interaction effects with
330 a mean PVE of 7.1 % (range = 5.1 – 10.1; Table 1). The MtCIM analysis of all
331 traits identified three significant and three almost significant (within one LOD
332 of the permutation threshold of 14.82 LOD) pleiotropic loci with multiple trait
333 effects (Table 2 and S5) [**see SUPPORTING INFORMATION**]. These

potential pleiotropic loci overlapped with the 2-LOD intervals of 14/29 of the individual trait QTLs, with up to four traits affected at each site (Table 1). Thus, up to 14 QTLs for a total of eight traitsexhibited pleiotropic effects. The "sampling without replacement" method using the 16.5 cM interval size found four trait pairs: auricle width and pollen number, capitulum number and node length, capitulum number and flowering time, and node length and leaf dissection, that showed significantly coincident QTL locations (Table S6) [**see SUPPORTING INFORMATION**]. Sampling without replacement analyses using a range of shorter interval sizes found similar evidence for coincident QTL locations, but failed to find any previously identified TRDLs that were significantly coincident with trait QTLs (Table S6) [**see SUPPORTING INFORMATION**].

Genetic diversity analysis

Both species exhibited similar levels of genetic diversity, with the highest diversity recorded for anonymous SSRs, followed in turn by EST-SSRs, EST-indels, and AFLPs, and other dominant markers (Tables S7 and S8) [**see SUPPORTING INFORMATION**]. Overall, inbreeding coefficients were not significantly different from zero in either species indicative of random mating (F_{IS} = 0.02 and 0.06 in *S. aethnensis* and *S. chrysanthemifolius*, respectively, Table S7) [**see SUPPORTING INFORMATION**]. The two species were significantly genetically differentiated across all marker types with overall F_{ST} of 0.28 and 0.31 observed for dominant markers and codominant markers, respectively (Tables S7 and S8) [**see SUPPORTING INFORMATION**].

Bayesian analyses of species differentiation showed that 4.7% of codominant markers, but 0% of dominant markers, were divergent outliers and that the same percentages of each marker type were significantly convergent outliers (Table 3). When population information was included in these analyses, the tests were more sensitive and identified 7.8% and 5.3% of significantly divergent codominant and dominant markers, respectively, and 4.7% and 0.8% of significantly convergent codominant and dominant markers, respectively (Figure 1, Table 3 and S9) [**see SUPPORTING INFORMATION**]. Significant outlier loci were distributed across most linkage groups of the genetic map (Figure 1) and showed no evidence of clustering according to a one-way binomial test of an excess of neighbouring pairs of outlier markers (p = 0.1754). However, significant negative associations between measures of species differentiation for marker loci and the genetic map distance from the nearest QTL peak were present (Figure 2). Similarly, there was evidence for negative associations between marker gene differentiation and genetic map distance from the nearest TRDL (Figure 2). A significant negative association between genetic differentiation and low recombination in the form of genetic map distance to closest neighbouring mapped locus was also found (Figure 2). Also, significant negative associations were present between genetic differentiation between species and the various intraspecific genetic diversity measures (Figure 3). In general, all of these associations were stronger for codominant than for dominant markers.

DISCUSSION

QTL architecture

Quantitative trait locus analysis identified one to five QTLs per trait and up to 29 QTLs in total for the 13 independent traits examined that distinguish the two *Senecio* species. In addition to resolving the primary effects of individual QTLs, MIM and MtCIM analyses provided evidence for epistatic interactions between four pairs of QTLs and possible pleiotropic effects at six loci affecting eight traits (Tables 1 and 2). Sampling without replacement tests indicated that QTL map locations were significantly clustered across the genetic map, with significant physical associations evident for four trait pairs (Table S5). Chapman *et al.* (2015) reported similar clustering of QTLs for species differences in an independent mapping study of *S. aethnensis* and *S. chrysanthemifolius*. However, their study did not investigate patterns of epistasis and pleiotropy. Regardless that the observed interactions between QTLs are due to epistasis or pleiotropy or physical linkage, they indicate that different traits are not genetically independent and that divergent selection acting on one trait would therefore also affect other traits.

A QTL architecture involving extensive physical and epistatic interactions between QTLs, together with pleiotropic effects of individual QTLs, should limit introgression between the two *Senecio* species on Mount Etna since hybridization would tend to break up gene complexes that control the expression of adaptive phenotypes in each species (Fenster *et al.* 1997; Turelli *et al.* 2001). The complex genomic architecture of interspecific divergence revealed in *Senecio* might reflect the evolutionary outcome of selection for non-independence of different QTLs controlling traits under divergent selection (Kirkpatrick and Barton 2006; Nosil *et al.* 2009; Nosil and Feder 2012; Yeaman 2013). Limiting recombination seems to be the crucial

factor permitting interacting QTLs to evolve into divergent co-adapted QTL complexes in the presence of gene flow. This can be achieved either through chromosomal rearrangement that causes recombination between rearranged regions to become deleterious (Feder *et al.* 2003; Lowry and Willis 2010; Twyford and Friedman 2015) or by evolution towards increased physical proximity (coincidence) through locally biased persistence, establishment, or translocation of QTLs (Via and West 2008; Nosil *et al.* 2009; Yeaman 2013). Genetic mapping indicates that *S. aethnensis* and *S. chrysanthemifolius* are not distinguished by major genome rearrangements (Brennan *et al.* 2014), which instead emphasizes the importance of QTL coincidence for this system (our results and those of Chapman *et al.*, 2015).

While TRDLs were not significantly coincident with QTLs for any trait according to the "sampling without replacement" method, a QTL affecting pollen viability co-located with a TRDL of large effect in linkage group AC1 (Figure 1). This finding is of interest as it adds to the result previously reported by Chapman et al (2015) of co-localization of TRDLs with QTLs affecting F_2 hybrid necrosis. Hybrid incompatibilities, such as decreased F2 pollen viability and hybrid necrosis, and their associated TRDLs, are expected to limit introgression across large genomic regions allowing further divergence of these regions during speciation (Barton and Bengtsson 1986; Barton and de Cara 2009).

Non-random patterns of divergence across the genome

Levels of molecular genetic diversity were similar in wild samples of both *Senecio* species, while genetic differentiation between species was moderate (F_{ST} = 0.31). Genetic diversity decreased from estimates based on

anonymous SSRs to EST-SSRs, to EST indels to AFLPs, corresponding to the expected ability of each marker type to resolve allelic variation (Tables S6 and S7) [**see SUPPORTING INFORMATION**]. Low levels of genetic differentiation between the two species were also reported by Muir *et al.* (2013), Osborne *et al.* (2013) and Chapman *et al.* (2013), based on surveys of microsatellite and sequence variation. We identified a small percentage of loci that were either significantly divergent or convergent (up to 7.8%) between species, dependent on the marker set analysed (Table 3 and S9) [**see SUPPORTING INFORMATION**]. This value is slightly greater than the 2.25% of outliers from a study of 8,854 loci recorded by Chapman *et al.* (2013) based on a comparison of the two species' transcriptomes, but the two findings are probably within the bounds of error given the different numbers of loci examined. More discussion about the functions of significantly divergent or convergent loci is provided in the supporting information. Inevitably, the 196 marker loci for which patterns of differentiation were compared to detect significant divergence between species in the present study provide only a very coarse-grained perspective across the whole genome, and many of the true genetic targets of selection will not have been surveyed.

Reduced effective gene flow in the vicinity of selected loci is often used to explain significantly differentiated loci and "islands of divergence" (Wu 2001; Feder and Nosil 2010). In support of this hypothesis, significant associations were found between interspecific genetic differentiation and genetic map distance to QTLs and TRDLs (Figure 2). These associations were negative with more highly differentiated loci positioned closer to QTLs or TRDLs. These results support previous findings that selection against

hybridization is important for maintaining species distinctiveness across the *Senecio* hybrid zone on Mount Etna (Brennan *et al.* 2009; Chapman *et al.* 2013; Chapman *et al.* 2015). However, independently of gene flow, within-species directional selection can also generate the same pattern of divergence via species-specific reductions in diversity (Cruikshank and Hahn 2014). The latter is amplified when it occurs in regions of low recombination as it causes longer genomic regions to be affected by selection at linked markers. In accordance with these hypotheses and in agreement with the findings of Chapman *et al.* (2015), we also found evidence for intraspecific selection in the form of significant negative associations between interspecific differentiation and local recombination, and between interspecific differentiation and intraspecific genetic diversity (Figures 2 and 3). It is plausible that *S. aethnensis* and *S. chrysanthemifolius* experience distinct localized selection pressures related to the very different environments they occupy at different elevations on Mount Etna. Such within-species selection would be expected to reduce within-species genetic diversity in the genomic regions experiencing selection. These findings therefore suggest a role for environment-specific extrinsic selection in maintaining the cline with elevation on Mount Etna. While this pattern of diversity might also signal past periods of isolation facilitating divergence , other genetic studies suggest that gene flow between the two species has probably been continuous throughout their history (Chapman *et al.* 2013, Osborne *et al.* 2013, Filatov *et al.* 2016).

CONCLUSIONS

Our study shows that phenotypic divergence across the elevational gradient on Mount Etna involves divergence of multiple quantitative traits

controlled by numerous interacting genes (QTLs). A breakdown in the complex genetic architecture of these traits following hybridization would be expected to reduce the fitness of most hybrid offspring and therefore contribute to introgression barriers between the two *Senecio* species. Our combined analyses of genetic differentiation, QTLs and TRDLs emphasize that divergence is non-randomly distributed across the genomes of these species and that both selection against hybrids between species and locally maladapted individuals within-species will act to maintain phenotypic divergence between the two species in the face of gene flow.

SUPPORTING INFORMATION

Genotype and genetic map data for the mapping family are available from Dryad at doi:10.5061/dryad.7b56k, while phenotype data for the mapping family and genotype and phenotype data for wild sampled individuals are available from Dryad at doi:10.5061/dryad.n3r2s.

The following **[SUPPORTING INFORMATION]** is available in the online version of this article:

File 1. Table S1. Information on wild sampled populations of *S. aethnensis* and *S. chrysanthemifolius*.

Table S2. Summary quantitative trait results for *S. aethnensis*, *S. chrysanthemifolius*, and a reciprocal F_2 *S. aethnensis* and *S. chrysanthemifolius* mapping family.

Table S3. Paired trait correlations in: (a) F_2AC progeny, (b) *Senecio aethnensis*, (c) *S. chrysanthemifolius*, and (d) all three samples.

Table S4. Comparison of summary quantitative trait locus results for a composite interval mapping (CIM) and multiple interval mapping (MIM)

analysis of a reciprocal F_2 *S. aethnensis* and *S. chrysanthemifolius* mapping family.

Table S5. Summary quantitative trait loci (QTLs) results from a multiple trait composite interval mapping (MtCIM) analysis compared to single trait QTL analyses of a reciprocal F_2 *S. aethnensis* and *S. chrysanthemifolius* mapping family.

Table S6. (a) "Sampling without replacement" test results for paired-trait QTL coincidence, (b) permutation tests of overall paired-trait QTL coincidence using different QTL and transmission ratio distortion loci (TRDL) datasets and genetic map interval sizes.

Table S7. Summary population genetic statistics for AFLPs and other dominant scored molecular genetic markers from *S. aethnensis* and *S. chrysanthemifolius* samples.

Table S8. Summary population genetic statistics for codominantly scored molecular genetic markers from *S. aethnensis* and *S. chrysanthemifolius* samples.

Table S9. Expressed sequence tag (EST) loci showing evidence for divergent or convergent selection between *S. aethnensis* and *S. chrysanthemifolius*.

File 2. Figure S1. Boxplots summarizing quantitative trait results for *S. aethnensis*, *S. chrysanthemifolius*, and a reciprocal F_2 mapping family that were included in the quantitative trait locus analysis.

Trait numbers in title correspond to the trait numbering system of Table 1. Bold horizontal lines indicate median values. Boxes indicate 25 to 75 percentile range. Lines indicate the range of values within 1.5 times the upper and lower quartiles, respectively. Points indicate values more extreme than

1.5 times the upper and lower quartiles. Asterisks indicate the trait values of the mapping family parents. No mapping family parental values were available for flowering time as these individuals were vegetatively propagated for comparison with their progeny.

File 3. Figure S2. Genetic map of a reciprocal F_2 *S. aethnensis* and *S. chrysanthemifolius* mapping family showing quantitative trait loci identified by composite interval mapping and marker loci that were significantly divergent or convergent between species.

Map distances in Kosambi centiMorgans are shown in the scale bar to the left of linkage groups. Linkage groups are represented by vertical bars with mapped locus positions indicated with horizontal lines. Weakly linked linkage groups (< 4 LOD or > 20 cM) that probably belong to the same chromosome are aligned vertically. Grey shading on linkage groups indicates regions exhibiting significant transmission ratio distortion (TRDLs). Locus names are listed to the left of linkage groups and mapped QTLs are listed to the right. “c” or “d” listed to the left of locus names indicates if that locus was identified as significantly convergent or divergent based on genetic differentiation analysis across sample populations; while > symbol to the left of locus names indicates if the locus was included in QTL analysis. QTLs were identified by composite interval mapping with significance determined if the LOD score exceeded the 0.95 quantile of 1000 data permutations. QTLs 2-LOD interval ranges are indicated with vertical lines with a bold horizontal line indicating the highest LOD score position. QTL summary information includes; trait names, “a” or “d” each followed by “+” or “-“ indicating additive or dominance effects and their

direction of effect supporting or opposing the observed species difference respectively, and the percent mapping family variance explained.

File 4. Additional text describing the genetic map, transmission ratio distortion analysis. composite interval mapping, QTL sign tests, and genetic diversity analyses.

SOURCES OF FUNDING

The research was funded by a NERC Grant NE/D014166/1 to R.J.A. as Principal Investigator. A.C.B. was supported during part of the writing of this paper by funding from FP7-REGPOT 2010-1, Grant No. 264125 EcoGenes.

CONTRIBUTIONS BY AUTHORS

R.J.A., A.C.B., and S.J.H. designed the research. A.C.B. performed the experiments and analysis. A.C.B. wrote the first draft and A.C.B, R.J.A., and S.J.H. contributed to revisions. The authors confirm that they have no conflicts of interest.

CONFLICTS OF INTEREST

No conflicts of interest.

ACKNOWLEDGEMENTS

We thank David Forbes for technical assistance and Ai-Lan Wang for help with measurement of quantitative traits. We thank the managing editor, Diana Wolf and anonymous reviewers for their suggestions to improve earlier versions of this paper.

LITERATURE CITED

579 Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierrne N, Boughman
580 J, Brelsford A, Buerkle A, Buggs R *et al.* 2013. Hybridization and speciation.
581 *Journal of Evolutionary Biology* 26: 229-246.
582 Abbott RJ, Brennan AC. 2014. Altitudinal gradients, plant hybrid zones and
583 evolutionary novelty. *Philosophical Transactions of the Royal Society Series B*
584 369: 20130346.
585 Barton NH, Bengtsson BO. 1986. The barrier to genetic exchange between
586 hybridising populations. *Heredity* 56: 357–376.
587 Barton NH, de Cara MAR. 2009. The evolution of strong reproductive
588 isolation. *Evolution* 63: 1171–1190.
589 Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling
590 hypothesis: why genome scans may fail to map local adaptation genes.
591 *Molecular Ecology* 20: 2044-2072.
592 Bouck A, Wessler SR, Arnold ML. 2007. QTL analysis of floral traits in
593 Louisiana Iris hybrids. *Evolution* 61: 2308-2319.
594 Breitling R, Li Y, Tesson BM, Fu J, Wu C, *et al.* 2008. Genetical genomics:
595 spotlight on QTL hotspots. *PLoS Genetics* 4: e1000232.
596 Brennan AC, Bridle JR, Wang A-L, Hiscock SJ, Abbott RJ. 2009. Adaptation
597 and selection in the *Senecio* (Asteraceae) hybrid zone on Mount Etna, Sicily.
598 *New Phytologist* 183: 702-717.
599 Brennan AC, Hiscock SJ, Abbott RJ. 2014. Interspecific crossing and genetic
600 mapping reveal intrinsic genomic incompatibility between two *Senecio* species
601 that form a hybrid zone on Mount Etna, Sicily. *Heredity* 113: 195-204.

602 Chapman MA, Hiscock SJ, Filatov DA. 2013. Genomic divergence during
603 speciation driven by adaptation to altitude. *Molecular Biology and Evolution*
604 30: 2553-2567.
605 Chapman MA, Hiscock SJ, Filatov DA. 2015. The genomic bases of
606 morphological divergence and reproductive isolation driven by ecological
607 speciation in *Senecio* (Asteraceae). *Journal Evolutionary Biology* 29: 98-113.
608 Churchill GA, Doerge RW. 1994. Empirical threshold values for quantitative
609 trait mapping. *Genetics* 138: 963-971.
610 Corbett-Detig RB, Zhou J, Clark, AG, Hartl DL, Ayroles JL. 2013. Genetic
611 incompatibilities are widespread within species. *Nature* 504: 135-137.
612 Coyne JA, Orr HA. 2004. *Speciation.* Sunderland, Massachusetts, USA:
613 Sinauer Associates.
614 Cruikshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of
615 speciation are due to reduced diversity, not reduced gene flow. *Molecular*
616 *Ecology* 23: 3133-3157.
617 Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of
618 programs to perform population genetics analyses under Linux and Windows.
619 *Molecular Ecology Resources* 10: 564-567.
620 Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating
621 genomic islands during ecological speciation. *Evolution* 64: 1729–1747.
622 Feder JL, Roethele JB, Filchak K, Niedbalski J, Romero-Severson J. 2003.
623 Evidence for inversion polymorphism related to sympatric host race formation
624 in the apple
625 maggot fly, *Rhagoletis pomonella. Genetics* 163: 939–953.

626 Fenster CB, Galloway LG, Chao L. 1997. Epistasis and its consequences for
627 the evolution of natural populations. *Trends in Ecology and Evolution* 12: 282–
628 286.
629 Filatov D, Osborne O, Papadopulos A. 2016. Demographic history of
630 speciation in *Senecio* altitudinal hybrid zone on Mt. Etna. *Molecular Ecology*
631 accepted.
632 Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci
633 appropriate for both dominant and codominant markers: a Bayesian
634 perspective. *Genetics* 180: 977-998.
635 Gagnaire PA, Normandeau E, Pavey SA, Bernatchez L. 2013. Mapping
636 phenotypic, expression and transmission ratio distortion QTL using RAD
637 markers in the Lake Whitefish (*Coregonus clupeaformis*). *Molecular Ecology*
638 22:3036-48.
639 Gompert Z, Parchman TL, Buerkle CA. 2012. Genomics of isolation in
640 hybrids. *Philosophical Transactions of the Royal Society Series B* 367: 439–
641 450.
642 Hatfield T, Schluter D. 1999. Ecological speciation in sticklebacks:
643 environment-dependent hybrid fitness. *Evolution* 53: 866–873.
644 James JK, Abbott RJ. 2005. Recent, allopatric, homoploid hybrid speciation:
645 the origin of *Senecio squalidus* (Asteraceae), in the British Isles from a hybrid
646 zone on Mount Etna, Sicily. *Evolution* 59: 2533-2547.
647 Jiang C, Zeng Z-B. 1995. Multiple trait analysis of genetic mapping for
648 quantitative trait loci. *Genetics* 140: 1111-1127.

649 Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E *et al.* 2012. The
650 genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484:
651 55–61.
652 Kalinowsky ST. 2005. HP-RARE 1.0: a computer program for performing
653 rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187-
654 189.
655 Kao C-H, Zeng Z-B. 1999. Multiple interval mapping for quantitative trait loci.
656 *Genetics* 152: 1203-1216.
657 Kirkpatrick M, Barton N. 2006 Chromosome inversions, local adaptation and
658 speciation. *Genetics* 173:419–434.
659 Lexer C, Rosenthal DM, Raymond O, Donovan LA, Rieseberg LH. 2005.
660 Genetics of species differences in wild annual sunflowers, *Helianthus annuus*
661 and *H. petiolaris*. *Genetics* 169: 2225–2239.
662 Lincoln SE, Daly MJ, Lander ES. 1993. *Constructing genetic linkage maps*
663 *with MAPMAKER/EXP version 3.0: A tutorial and reference manual.*
664 Cambridge, Massachusetts, USA: Whitehead Institute for Biomedical
665 Research Technical Report 3rd edition.
666 Lindtke D, Buerkle CA. 2015. The genetic architecture of hybrid
667 incompatibilities and their effect on barriers to introgression in secondary
668 contact. *Evolution* 69: 1987–2004.
669 Lowry DB, Willis JH. 2010. A widespread chromosomal inversion
670 polymorphism
671 contributes to a major life-history transition, local adaptation, and reproductive
672 isolation. PLoS Biol 8: e1000500.

673 Muir G, Osborne OG, Sarasa J, Hiscock SJ, Filatov DA. 2013. Recent
674 ecological selection on regulatory divergence is shaping clinal variation in
675 *Senecio* on Mount Etna. *Evolution* 67: 3032-3042.
676 NCBI Resource Coordinators. 2013. Database resources of the National
677 Center for Biotechnology Information. *Nucleic Acids Research* 41: D8-D20
678 Nosil P. 2012. *Ecological speciation*. Oxford, UK: Oxford University Press.
679 Nosil P, Feder JL. 2012. Genomic divergence during speciation: causes and
680 consequences. *Philosophical Transactions of the Royal Society Series B* 367:
681 332-342.
682 Nosil P, Funk DJ, Ortíz-Barrientos D. 2009. Divergent selection and
683 heterogeneous genomic divergence. *Molecular Ecology* 18: 375–402.
684 Orr HA. 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics*
685 144: 1331–1335.
686 Orr HA, Turelli M. 2001. The evolution of postzygotic isolation: accumulating
687 Dobzhansky–Muller incompatibilities. *Evolution* 55: 1085–1094.
688 Osborne OG, Batstone TE, Hiscock SJ, Filatov DA. 2013. Rapid speciation
689 with gene flow following the formation of Mt. Etna. *Genome Biology and*
690 *Evolution* 5: 1704-1715.
691 Paterson AH. 2002. What has QTL mapping taught us about plant
692 domestication? *New Phytologist* 154: 591-608.
693 Peakall R, Smouse PE. 2006. GENALEX 6: genetic analysis in Excel.
694 Population genetic software for teaching and research. *Molecular Ecology*
695 *Notes* 6: 288-295.
696 R Development Core Team. 2011. *R: A language and environment for*
697 *statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

698 Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM,
699 Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T,
700 Durphy JL, Schwarzbach AE, Donovan LA, Lexer C. 2003. Major ecological
701 transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211–
702 1216.
703 Rogers SM, Bernatchez L. 2007. The genetic architecture of ecological
704 speciation and the association with signatures of selection in natural lake
705 whitefish (*Coregonus* sp. Salmonidae). *Molecular Biology and Evolution* 24:
706 1423–1438.
707 Rogers SM, Mee JA, Bowles E. 2013. The consequences of genomic
708 architecture on ecological speciation in postglacial fishes. *Current Zoology* 59:
709 53-71.
710 Servedio MR, Van Doorn GS, Kopp M, Frame AM, Nosil P. 2011. Magic traits
711 in speciation: 'magic' but not rare? *Trends in Ecology and Evolution* 26: 389–
712 397.
713 Smadja CM, Butlin RK. 2011. A framework for comparing processes of
714 speciation in the presence of gene flow. *Molecular Ecology* 20: 5123–5140.
715 Stinchcombe JR, Hoekstra HE. 2008. Combining population genomics and
716 quantitative genetics: finding the genes underlying ecologically important
717 traits. *Heredity* 100: 158-170.
718 Strasburg JL, Sherman NA, Wright KM, Moyle LC, Willis JH, Rieseberg LH.
719 2012. What can patterns of differentiation across plant genomes tell us about
720 adaptation and speciation? *Philosophical Transactions of the Royal Society*
721 *Series B* 367: 364-373.

722 Taylor SJ, Rojas LD, Ho S-W, Martin NH. 2012. Genomic collinearity and the
723 genetic architecture of floral differences between the homoploid hybrid
724 species *Iris nelsonii* and one of its progenitors, *Iris hexagona*. *Heredity* 110:
725 63-70.
726 Turelli M, Barton NH, Coyne JA. 2001. Theory and speciation. *Trends in*
727 *Ecology and Evolution* 19: 490–496.
728 Turner T, Hahn M, Nuzhdin SV. 2005. Genomic islands of speciation in
729 *Anopheles gambiae*. *PLoS Biology*, 3, e285.
730 Twyford AD, Friedman J. 2015. Adaptive divergence in the monkey flower
731 *Mimulus guttatus* is maintained by a chromosomal inversion. *Evolution* 69:
732 1476-1486.
733 Via S, West J. 2008. The genetic mosaic suggests a new role for hitchhiking
734 in ecological speciation. *Molecular Ecology* 17: 4334–4345.
735 Wang S, Basten CJ, Zeng, Z-B. 2011. *Windows QTL Cartographer 2.5.*
736 Raleigh, North Carolina, USA: Department of Statistics, North Carolina State
737 University.
738 Whiteley AR, Derome N, Rogers SM, St-Cyr J, *et al.* 2008. The phenomics
739 and expression quantitative trait locus mapping of brain transcriptomes
740 regulating adaptive divergence in Lake Whitefish species pairs (*Coregonus*
741 sp.). *Genetics* 180: 147–164.
742 Wu C-I. 2001. The genic view of the process of speciation. *Journal of*
743 *Evolutionary Biology* 14: 851–865.
744 Yeaman S. 2013. Genomic rearrangements and the evolution of clusters of
745 locally adaptive loci. *Proceedings of the National Academy of Sciences* 110:
746 E1743-E1751.

747 Yeaman S, Whitlock MC. 2011. The genetic architecture of adaptation under
748 migration-selection balance. *Evolution* 65:1897–1911.
749

FIGURE LEGENDS

Figure 1. Genetic map of a reciprocal F_2 *S. aethnensis* and *S. chrysanthemifolius* mapping family showing quantitative trait loci (QTLs) identified by multiple interval mapping and marker loci that were significantly divergent or convergent between species. Map distances in Kosambi centiMorgans are shown in the scale bar to the left of linkage groups. Linkage groups are represented by vertical bars with mapped locus positions indicated with horizontal lines. Weakly linked linkage groups (< 4 LOD or > 20 cM) that probably belong to the same chromosome are aligned vertically. Grey shading on linkage groups indicates regions exhibiting significant transmission ratio distortion (TRDLs). Locus names are listed to the left of linkage groups and mapped QTLs are listed to the right. “c” or “d” listed to the left of locus names indicates if locus was identified to be significantly convergent or divergent based on genetic differentiation analysis across sample populations; while the ‘>’ symbol to the left of locus names indicates if the locus was included in QTL analysis. QTLs were identified by multiple interval mapping with significance testing by Bayesian information criterion model comparisons. QTLs 2-LOD ranges are indicated with vertical lines with a bold horizontal line indicating the highest LOD score position. QTL summary information includes: trait names, “a” or “d” each followed by “+” or “-“ indicating additive or dominance effects and their direction of effect supporting or opposing the observed species difference respectively, and the percent mapping family variance explained (PVE).

Figure 2. Relationships between genetic differentiation and genetic map distance to: the nearest quantitative trait locus (QTL) peak, the nearest

775 transmission ratio distortion locus (TRDL), or the nearest mapped marker.
776 Presented *p* values summarize Spearman rank correlation tests. All significant
777 associations were negative. Sample sizes were 48 codominant loci and 63
778 dominant loci. Loci on linkage groups without a QTL or TRDL peak were
779 assigned an unlinked genetic map distance of 50 cM.
780 **Figure 3.** Relationships between genetic differentiation and genetic diversity
781 of wild sampled *S. aethnensis* and *S. chrysanthemifolius*. Presented *p* values
782 summarize Spearman rank correlation tests. All significant associations were
783 negative. Sample sizes were 65 codominant loci and 145 dominant loci.